Raymond Salzmann Machine Learning Homework 1

10/11/2020

Problem 1:

$r_{ij} = \{-1, 1\}$

$u_i \in \mathbb{R}^d \; ; \; v_j \in \mathbb{R}^d \; ; \; i = 1, \dots, N \; ; \; j = 1, \dots, M$

- Data: $\{(u_i, v_j)\}_{i,j}^{N,M}$
- Model: $r_{ij} \mid U, V \overset{ind}{\sim} \text{Bernoulli}(\Phi(u_i^T v_j / \sigma))$ for all $(i, j) \in \Omega$
- Priors: $u_i \overset{iid}{\sim} N(0, cI) \quad ; \quad v_j \overset{iid}{\sim} N(0, cI)$

$$\ln P(R, u, v) = \underbrace{\int q(\phi) \ln \frac{p(R, u, v, \phi)}{q(\phi)} d\phi}_{\mathcal{L}(u,v)} + \int q(\phi) \ln \frac{q(\phi)}{p(\phi \mid R, u, v)} d\phi$$

$\phi = \{\phi_{ij}\}$ for $(i, j) \in \Omega$ and $r_{ij} = \text{sign}(\phi_{ij})$, $\phi_{ij} \sim N(u_i^T v_j, \sigma^2)$

Part A:

$X = U \qquad w = V$

From the EM algorithm we know that we need to select q such that:

$$q(\phi) = p(\phi \mid R, u, v)$$

Thus:

$$p(\phi \mid R, u, v) = \frac{\prod_{(i,j)} p(r_{ij} \mid \phi_{ij}, v_j, u_i)\, p(\phi_{ij} \mid v_j, u_i)}{\prod_{(i,j)} \int p(r_{ij} \mid \phi_{ij}, v_j, u_i)\, p(\phi_{ij} \mid v_j, u_i) d\phi_{ij}}$$

$$= \prod_{(i,j)\in\Omega} \frac{p(r_{ij} \mid \phi_{ij}, v_j, u_i)\, p(\phi_{ij} \mid v_j, u_i)}{\int p(r_{ij} \mid \phi_{ij}, v_j, u_i)\, p(\phi_{ij} \mid v_j, u_i)\, d\phi_{ij}}$$

$$p(\phi \mid R, u, v) = \prod_{(i,j)\in\Omega} p(\phi_{ij} \mid r_{ij}, v_j, u_i)$$

$\therefore$ Since $q(\phi) = p(\phi \mid R, u, v)$

then $q(\phi) = \prod_{(i,j)\in\Omega} p(\phi_{ij} \mid r_{ij}, v_j, u_i)$

thus, $q(\phi) = \prod_{(i,j)\in\Omega} q(\phi_{ij})$

Consider:

$$r_{ij} = \mathbb{1}(\phi_{ij} > 0), \quad \phi_{ij} \sim N(u_i^T v_j, \sigma^2)$$

For $r_{ij} = 1$: $p(r_{ij} = 1, \phi_{ij} \mid v_j, u_i) = p(r_{ij} = 1 \mid \phi_{ij})\, p(\phi_{ij} \mid v_j, u_i)$

$$= \mathbb{1}(\phi_{ij} > 0)(2\pi\sigma^2)^{-\frac{1}{2}} e^{\frac{-1}{2\sigma^2}(\phi_{ij} - u_i^T v_j)^2}$$

For $r_{ij} = -1$: $p(r_{ij} = -1, \phi_{ij} \mid v_j, u_i) = \mathbb{1}(\phi_{ij} \leq 0)(2\pi\sigma^2)^{-\frac{1}{2}} e^{\frac{-1}{2\sigma^2}(\phi_{ij} - u_i^T v_j)^2}$

Next: $\int p(r_{ij} = 1, \phi_{ij} \mid v_j, u_i)\, d\phi_{ij} = \int_{-\infty}^{\infty} \mathbb{1}(\phi_{ij} > 0)(2\pi\sigma^2)^{-\frac{1}{2}} e^{\frac{-1}{2\sigma^2}(\phi_{ij} - u_i^T v_j)^2}$

$$= \int_{0}^{\infty} (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(\phi_{ij} - u_i^T v_j)^2}$$

$$= P(\phi_{ij} > 0)$$

Thus: $$p(\phi_{ij} \mid r_{ij}, v_j, u_i) = \frac{p(r_{ij} \mid \phi_{ij})\, p(\phi_{ij} \mid u_i, v_j)}{\int p(r_{ij} \mid \phi_{ij})\, p(\phi_{ij} \mid u_i, v_j)\, d\phi_{ij}}$$

$$= \frac{\mathbb{1}\{\text{sign}(\phi_{ij}) = 2r_{ij} - 1\} e^{\frac{-1}{2\sigma^2}(\phi_{ij} - u_i^T v_j)^2}}{\int_{-\infty}^{\infty} \mathbb{1}\{\text{sign}(\phi_{ij}) = 2r_{ij} - 1\} e^{\frac{-1}{2\sigma^2}(\phi_{ij} - u_i^T v_j)^2} d\phi_{ij}}$$

So, if $r_{ij} = 1$, $\phi_{ij} > 0$ and if $r_{ij} = -1$, $\phi_{ij} \leq 0$

$$\therefore \quad q(\phi_{ij}) = TN_{r_{ij}}(u_i^T v_j, \sigma^2)$$

And, $$q(\phi) = \prod_{(i,j)} q(\phi_{ij})$$

$$\therefore \quad q(\phi) = TN_{r_{ij}}(U^T V, \sigma^2)$$

Part B:

$$\mathcal{L}_t(u,v) = \int q_t(\phi) \ln \frac{p(U, V, R, \phi)}{q_t(\phi)} d\phi$$

$$= \int q_t(\phi) \ln p(U, V, R, \phi)\, d\phi - \int q_t(\phi) \ln(q_t(\phi))\, d\phi$$

$$\propto \int q_t(\phi) \ln p(U, V, R, \phi)\, d\phi$$

$$\propto \int q_t(\phi) \ln p(R, \phi \mid U, V)\, p(U, V)$$

$$\propto \int q_t(\phi)\left[\ln p(R,\phi \mid u,v) + \ln p(u,v)\right] d\phi$$

$$\propto \int q_t(\phi)\left[\ln \prod_{(i,j)} p(r_{ij},\phi_{ij} \mid u_i,v_j) + \ln p(u,v)\right] d\phi$$

$$\propto \int q_t(\phi)\left[\ln \prod_{(i,j)} p(r_{ij},\phi_{ij} \mid u_i,v_j) d\phi + \int q_t(\phi) \ln p(u,v) d\phi\right.$$

$$\propto \mathbb{E}_q\left[\sum_{i,j} \ln p(r_{ij},\phi_{ij} \mid v_j,u_i)\right] + \ln p(u,V)$$

$$\therefore \mathcal{L}(u,V) = \ln p(u,v) + \sum_{(i,j)} \mathbb{E}_q\left[\ln p(r_{ij},\phi_{ij} \mid V_j,u_i)\right] + \text{const.}$$

Since $$p(R,\phi,V \mid u) = p(V) \prod_{(i,j)} p(r_{ij} \mid \phi_{ij}) p(\phi_{ij} \mid u_i,v_j)$$

Thus we have that:

$$\mathcal{L}(u,V) = \frac{-1}{2c} V^T V - \frac{1}{2c} U^T u - \frac{1}{2\sigma^2} \mathbb{E}_q\left[(\phi_{ij} - u_i^T v_j)^2\right] + \text{const.}$$

$$\mathcal{L}(u,V) = \frac{-1}{2c} V^T V - \frac{1}{2c} u^T u$$

$$-\frac{1}{2\sigma^2} \sum_{(i,j)} \left(v_j^T u_i u_i^T v_j - 2 v_j^T u_i \mathbb{E}_q[\phi_i]\right) + \text{const.}$$

$$\mathcal{L}(u,V) = \frac{-1}{2c} \sum_{i=1}^{N} (u_i^T u_i) - \frac{1}{2c} \sum_{j=1}^{M} (v_j^T v_j)$$

$$-\frac{1}{2\sigma^2} \sum_{(i,j)} \left(v_j^T u_i u_i^T v_j - 2 v_j^T u_i \mathbb{E}_q[\phi_{ij}]\right) + \text{const.}$$

And: $\mathbb{E}_q[\phi_{ij}] = \begin{cases} u_i^T v_j + \sigma \times \dfrac{\Phi'(-u_i^T v_j/\sigma)}{1-\Phi(-u_i^T v_j/\sigma)} & \text{if } r_{ij}=1 \\ u_i^T v_j + \sigma \times \dfrac{-\Phi'(-u_i^T v_j/\sigma)}{\Phi(-u_i^T v_j/\sigma)} & \text{if } r_{ij}=-1 \end{cases}$

Part C:

$$\nabla_{u_i} \mathcal{L}(u,v) = \frac{-1}{c} u_i - 0 - \sum_{(i,j)\in\Omega} \frac{1}{2\sigma^2}\left(v_j^T u_i u_i^T v_j - 2 v_j u_i \mathbb{E}_q[\phi_{ij}]\right)$$

$$= \frac{-1}{c} u_i - \sum_{(i,j)\in\Omega} \frac{1}{2\sigma^2}\left(2 v_j^T u_i v_j - 2 v_j \mathbb{E}_q[\phi_{ij}]\right)$$

$$\frac{1}{c} u_i = \sum_{(i,j)\in\Omega} \frac{-1}{2\sigma^2}\left(2 v_j^T u_i v_j - v_j \mathbb{E}_q[\phi_{ij}]\right)$$

$$\frac{1}{c} u_i = -\sum_{(i,j)} \frac{v_j^T u_i v_j}{\sigma^2} + \sum_{(i,j)} \frac{v_j \mathbb{E}_q[\phi_{ij}]}{\sigma^2}$$

$$\frac{1}{c} u_i + \sum_{(i,j)} \frac{v_j^T u_i v_j}{\sigma^2} = \sum_{(i,j)} \frac{v_j \mathbb{E}_q[\phi_{ij}]}{\sigma^2}$$

$$u_i\left(\frac{1}{c} + \sum_{(i,j)} \frac{v_j^T v_j}{\sigma^2}\right) = \sum_{(i,j)} \frac{v_j \mathbb{E}_q[\phi_{ij}]}{\sigma^2}$$

$$u_i = \left(\frac{1}{c} + \sum_{(i,j)\in\Omega} \frac{v_j^T v_j}{\sigma^2}\right)^{-1}\left(\sum_{(i,j)\in\Omega} \frac{v_j \mathbb{E}_q[\phi_{ij}]}{\sigma^2}\right)$$

$$\nabla_{v_j} \mathcal{L}(U,V) = 0 - \frac{1}{c} v_j - \sum_{(i,j)} \frac{1}{2\sigma^2}\left(2 v_j u_i u_i^T - 2 u_i \mathbb{E}_q[\phi_{ij}]\right)$$

$$= -\frac{1}{c} v_j - \sum_{(i,j)} \frac{1}{2\sigma^2}\left(2 v_j u_i u_i^T - 2 u_i \mathbb{E}_q[\phi_{ij}]\right)$$

$$\frac{1}{c} v_j = \sum_{(i,j)} \frac{- v_j u_i u_i^T}{\sigma^2} + \sum_{(i,j)} \frac{u_i \mathbb{E}_q[\phi_{ij}]}{\sigma^2}$$

$$\frac{1}{c} v_j + \sum_{(i,j)} \frac{v_j u_i u_i^T}{\sigma^2} = \sum_{(i,j)} \frac{u_i \mathbb{E}_q[\phi_{ij}]}{\sigma^2}$$

$$v_j\left(\frac{1}{c} + \sum_{(i,j)} \frac{u_i u_i^T}{\sigma^2}\right) = \sum_{(i,j)} \frac{u_i \mathbb{E}_q[\phi_{ij}]}{\sigma^2}$$

$$\boxed{v_j = \left(\frac{1}{c} + \sum_{(i,j)\in\Omega} \frac{u_i u_i^T}{\sigma^2}\right)^{-1}\left(\sum_{(i,j)\in\Omega} \frac{u_i \mathbb{E}_q[\phi_{ij}]}{\sigma^2}\right)}$$

Initialize u_i and v_j, for our case $u_i \overset{iid}{\sim} N(0, cI)$ and $v_j \overset{iid}{\sim} N(0, cI)$

Part D:

(1) Initialize the unknown model variables to be used in the algorithm, in our case that is to initialize u_i and v_j. While normally the unknown variables would be initialized to zero, we will intialize ours to $u_i \overset{iid}{\sim} N(0, cI)$, $v_j \overset{iid}{\sim} N(0, cI)$ as defined by their priors.

(2) For the number of iterations you would like to run the algorithm for, loop the following steps:

(a) First we must calculate the Expectation $\mathbb{E}_q[\phi_i]$ is under the conditional posterior, which we derived in Question 1(a) above. This is known as the E-step, or the Expectation Step. In our model,

$$\mathbb{E}_q[\phi_{ij}] = \begin{cases} u_i^T v_j + \sigma \times \dfrac{\Phi'(-u_i^T v_j/\sigma)}{1-\Phi(-u_i^T v_j/\sigma)} & \text{if } r_{ij}=1 \\ \\ u_i^T v_j + \sigma \times \dfrac{-\Phi'(-u_i^T v_j/\sigma)}{\Phi(-u_i^T v_j/\sigma)} \end{cases}$$

(b) We must now update the unknown model vectors using the above expectations. In our case we found a u_i and v_j that we could generalize to the entirety of the variables U and V. In order to do this we took the gradiet of $\mathcal{L}(U,V)$ formally defined as $\int q(\phi) \ln \dfrac{P(R,U,V,\phi)}{q(\phi)} = \mathcal{L}(U,V)$. By taking the gradient of $\mathcal{L}(U,V)$ with respect to each of the unknown model variables U and V. This is called the M-step, or the Maximization Step.

(c) For the last step, we must calculate the log joint likelihood. We use the updated U and V values that were found in the M-step, based on the $\mathbb{E}_q[\phi_i]$, which is updated at the start of each loop. In our case this can be defined as

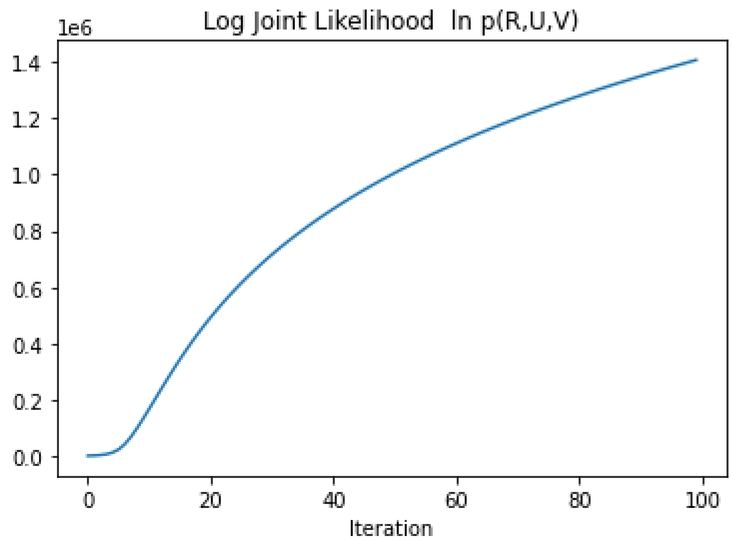
$$\ln p(R,U,V) = \sum_i \left[\frac{d}{2}\ln\left(\frac{1}{2\pi c}\right) - \frac{1}{2c}(u_i^T u_i)\right]$$

$$+ \sum_j \left[\frac{d}{2}\ln\left(\frac{1}{2\pi c}\right) - \frac{1}{2c}(v_j^T v_j)\right)$$

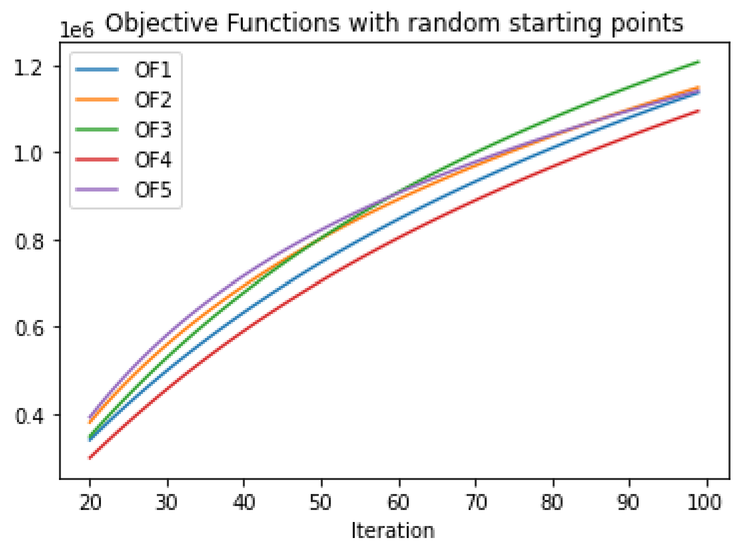
$$+ \sum_{(i,j)} \left[r_{ij}\ln\left[\Phi(u_i^T v_j/\sigma)\right]\right] + \sum_{(i,j)}\left[(1-r_{ij})\ln\left(1-\Phi\left(u_i^T v_j/\sigma\right)\right)\right]$$

Problem 2: Code in Question2.py

Q2 Part B:



Q2 Part B:



Q2 Part C:

Confusion Martix for EM Algorithm	1	-1
1	1431	835
-1	608	2126

Problem 3: Code is in Question3.py

- Data is from Problem 2
- Treat r_{ij} as a real-valued observation
- $d=5$, $c=1$, $\sigma^2=1$

Part A:

$$\ln p(R,u,v) = \ln[p(R|u,v)\,p(u,v)]$$

$$= \ln[p(R|u,v)\,p(u)\,p(v)]$$

$$\therefore \ln p(R,u,v) = \ln p(R|u,v) + \ln p(u) + \ln p(v)$$

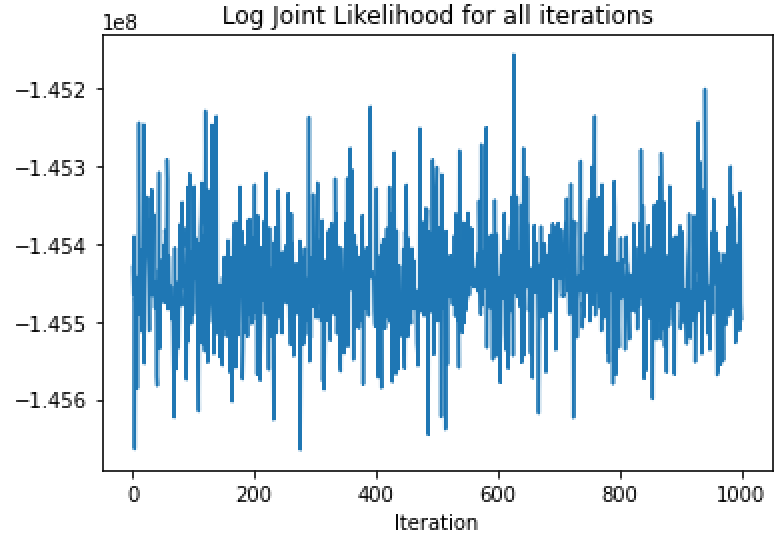
$$= \frac{-1}{2\sigma^2}\left(R^TR - RR^Tu^Tv + v^Tuu^Tv\right) - \frac{1}{2c}u^Tu - \frac{1}{2c}v^Tv + \text{const.}$$

We can now replace the full model variables with the sums of the individuals

$$\ln p(r_{ij}, u_i, v_j) = \sum_{(i,j)\in\Omega} \left[\frac{-1}{2\sigma^2}\left(r_{ij}^2 - 2r_{ij}u_iv_j + v_j^Tu_iu_i^Tv_j\right) - \frac{1}{2c}\left(u_i^Tu_i + v_j^Tv_j\right)\right] + \text{const.}$$

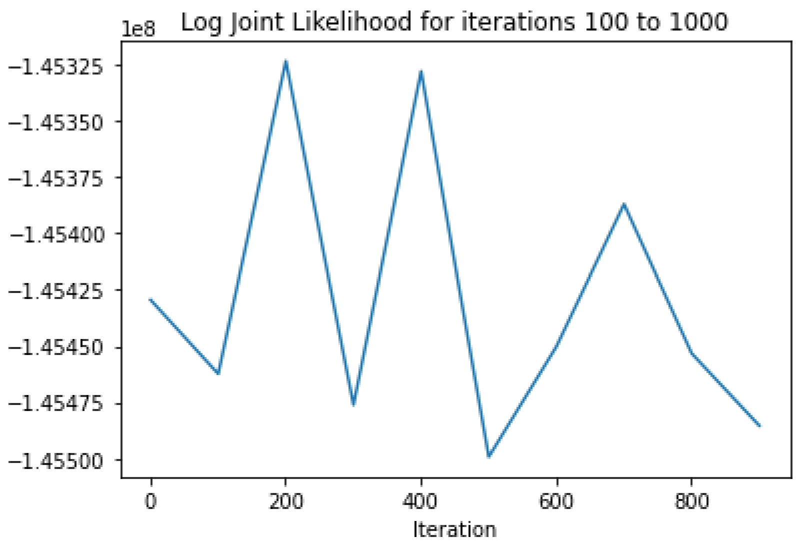
Q3

Part B:



Q3

Part B:



Q3

Part C:

Confusion Matrix for Gibbs Sampling	1	-1
1	1146	1120
-1	608	2126